DOCUMENT RESUME

ED 186 439 TM 800 005

AUTHOR Kern, Jeffrey M.; MacDonald, Marian L.
TITLE Assessing Assertion: An Investigation of Construct Validity and Reliability.
PUB DATE 80
NOTE 33p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Anxiety; *Assertiveness; Females; Higher Education; Observation; Predictive Measurement; *Psychological Testing; Responses; Self Evaluation (Individuals); *Situational Tests; *Test Reliability; *Test Validity
IDENTIFIERS Autonomic Perception Questionnaire; College Self Expression Scale; *College Womens Assertion Sample; Conflict Resolution Inventory; Fear of Negative Evaluation Scale (Watson Friend); Multitrait Multimethod Techniques; Social Reaction Inventory (Richardson Tasto); S R Inventory of Anxiousness (Endler et al); Timed Behavior Checklist

ABSTRACT

The reliability and meaning of assertiveness tests were explored using 120 female undergraduates. Several self-report inventories (the College Self-Expression Scale, Conflict Resolution Inventory, and a global rating from one to seven) were administered, as were three anxiety measures (Timed Behavior Checklist, response latency, and response duration) which were collected during administration of the College Women's Assertion Sample (CWAS). The CWAS involved 50 audiotaped interpersonal situations, during which the subjects' role-playing responses were observed. Five self-report measures of anxiety (the Fear of Negative Evaluation Scale, S-R Inventory of Anxiousness, Social Reaction Inventory, Autonomic Perception Questionnaire, and a global rating of anxiousness from one to seven) were then administered. Retesting at one-, four-, or ten-week intervals supported the reliability of the behavioral sampling and the Conflict Resolution Inventory. Convergent validity was evident for all the assertion tests, but only the CWAS showed strong discriminant validity. The global self-ratings were minimally useful, and the behavioral sampling successfully differentiated between aggressive, assertive, submissive, and aggressive-submissive subjects. The use of multiple measures was recommended. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED186439

Assessing Assertion: An Extended Report on an Investigation of Construct Validity and Reliability

Jeffrey M. Kern
Texas A & M University

Marian L. MacDonald
University of Massachusetts, Amherst

Running head: Assessing Assertion

Address for correspondence related to this manuscript:

Marian L. MacDonald, Ph.D.
Department of Psychology
University of Massachusetts
Amherst, Massachusetts 01003

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Macdonald

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM 800005

Abstract

The reliability and meaning of assertion measures was explored. Behavior sample and self-report indices of assertion and anxiety were collected over 3 retest intervals: 1 week (N=40), 4 weeks (N=40), and 10 weeks (N=40). Indices of temporal stability generally supported the reliability and sensitivity of the behavior sample measure and one of the self-report forms. Convergent validity was evident for all of the assertion instruments, but only the role-play measure showed pronounced discriminant validity. The utility of global self ratings on all counts was minimal. The behavior sample measure of assertion was the only variable which successfully differentiated between aggressive, assertive, submissive, and aggressive-submissive subject subtypes.

Assessing Assertion: An Investigation of Construct Validity and Reliability

The popularity of research involving assertion has been accompanied by the appearance of a number of different instruments designed to measure it. Amidst this instrument proliferation, regrettably few studies have been reported investigating relationships among the different measures (Hersen & Bellack, 1977). Furthermore, scant attention has been paid to the evaluation of various instruments' basic psychometric properties (Nay, 1977).

Knowledge gaps about instrument relationships and basic psychometric properties have several undesirable effects. First, unrecognized instrument differences, when they do exist, diminish the comparability of results across studies to such an extent that findings may be reported as contradictory when, in fact, they are not (compare, for example, Hersen, Eisler, Miller, Johnson, & Pinkston, 1973, with McFall & Twentyman, 1973). Second, uncertainty about relationships between instruments believed to measure the same thing contributes to a lack of clarity in the actual meaning of the construct being measured. And finally, difficulties in interpreting the substance of data from outcome studies arise when basic instrument properties such as temporal stability have not been evaluated independently for relevant dependent variables.

Some temporal stability data have been reported (e.g., Galassi, Delo, Galassi, & Bastein, 1974; Gambrill & Richey, 1-75). To date, data supporting the validities of various assertion measures have been derived primarily from demonstrations of their sensitivity to differential treatment effects (e.g., Eisler, Hersen, & Miller, 1973; Hersen, Eisler, & Miller, 1973; Kazdin, 1975; McFall & Lillesand, 1971; McFall & Marston, 1970) and demonstrations of their successfully distinguishing between two contrasted groups (e.g., Eisler, Miller, & Hersen, 1973; Galassi, Hollandsworth, Radecki, Gay, Howe, & Evans, 1976; Gambrill & Richey, 1975). While studies have addressed the convergent validities of several measures (e.g., Friedman, 1971; Galassi & Galassi, & Litz, 1974; Gambrill & Richey, 1975; McFall & Lillesand, 1971), disciminant validation within the context of Campbell and Fiske's (1959) multitrait-multimethod model, a rigorous methodology allowing evaluation of convergency and divergence given method factors, has not yet been applied. The present study was conceived, then, to apply the multitrait-multimethod strategy to several assertion instruments drawn from two measurement methods, behavior sample and self, report while studying the stabilities of these same instruments over time (see Kern & MacDonald, 1980).

Method

Subjects

One hundred and twenty undergraduate women participated in this study. Subjects were recruited on a voluntary basis and received research credit for their participation. They were predominantly freshpersons.

Procedure

Participants were introduced to the experiment and informed that in order to receive research credit they were required to return to the laboratory at a convenient future time. Furthermore, it was explained that this experiment involved two segments: completing several self-report measures, and responding to audiotaped stimuli while being observed through a one-way mirror.

Assenting subjects (120 of 120) were assigned randomly to one of three groups: a one-week, four-week, or ten-week test-retest interval. Within each group, experimental segments were completely counterbalanced across subjects. To control for additional order effects, the administration of the written measures, with the exception of the Autonomic Perception Questionnaire (Mandler, Mandler, & Uviller, 1958), was incompletely counterbalanced between subjects such that each measure occurred equally often at each ordinal position. The Autonomic Perception Questionnaire (Mandler et al., 1958)

was always administered immediately following the audiotaped segment.

Subjects were tested individually by one of eight trained examiners. Test order, instructions, and examiners were held constant for all subjects across both testing periods.

Measures

Behavior sample measure of assertion. The College Women's Assertion Sample (CWAS: MacDonald, 1974, 1978) is a role-play measure comprised of 50 audiotaped interpersonal situations empirically related to assertion. Subjects' verbatim responses are recorded by an examiner, who refrains from any interaction with the testee during testing. Response latencies and durations are also recorded. Examples of test situations are as follows:

(1) You're going by yourself to a concert. As you get to your seat, you see that someone else is sitting there. You ask him to check the seat number on his ticket, and he replies, "I did that when I sat down."

(2) Last week you bought a pocket calculator and resolved never to loan it since it cost so much. A very good friend calls and says, "I have a take-home statistics quiz I just can't finish by hand, and it's due tomorrow. Can I borrow your new calculator? I'll be really careful with it."

CWAS scoring involves classifying responses as instances of pre-established response levels ranging from highly aggressive to highly submissive and assigning classified responses scoring weights derived in prior studies with similar populations. Full details of administration and scoring procedures, as well as the construction of the CWAS, may be found in MacDonald (1974; 1978; Note 1).

Self-report measures of assertion. Subjects completed three self-report measures of assertion: The College Self-Expression Scale (CSES; Galassi, Delo, Galassi, & Bastien, 1974), the Conflict Resolution Inventory (CRI; McFall & Lillesand, 1971), and a global self-rating of "overall assertion level" on a seven-point, anchored scale.

Behavior sample measures of anxiety. Three behavior sample measures of anxiety were included; all were collected during the administration of the CWAS. The first was a modified, seven category version of the Timed Behavior Checklist (TBCL; Paul, 1966); categories included were: (1) foot/leg movement; (2) hand/arm movement exclusive of gestures; (3) body movement; (4) lip grimacing; (5) hand/arm restraint; (6) closed eyes, and (7) facial tension. With this measure, subjects were observed through a one-way mirror throughout the CWAS administration, a period of approximately 45 minutes. Baserate levels of observed anxiety were recorded for two 20-second intervals during the instructional prelude. Thereafter, to

allow for a comparison of overt anxiety behaviors before and during responding to assertion situations, TBCL observations occurred during half of the item presentations and half of the item responses. Order of the two was counterbalanced across subjects within groups. TBCL observational periods, then, were not strictly time contingent, but rather were event contingent, varying in length according to the fixed item and variable response durations. Rater reliability was assessed through random, unobtrusive spotchecks by a criterion rater (Taplin & Reed, 1973) through a second one-way mirror; there were differences in observational angels, then, which may have somewhat attenuated reliability estimates. The second and third behavior sample measures of anxiety were average latency and duration of response to the 50 items of the CWAS.

Self-report measures of anxiety. Five measures of self-reported anxiety were employed: (1) The Fear of Negative Evaluation Scale (FNE; Watson & Friend, 1969), (2) a modified, 10-item S-R Inventory of Anxiousness (SRIA; Endler, Hunt & Rosenstein, 1962), with the 10 items being those CWAS situations correlating most highly with total CWAS score in previous work (MacDonald, 1974, 1978, Note 2), (3) Scale II from the Social Reaction Inventory (SRI; Richardson & Tasto, 1976), (4) a global self-rating of "Overall level of anxiousness" on a seven-point, anchored

scale, and (5) the Autonomic Perception Questionnaire (APQ; Mandler et al., 1958), which was administered directly following the CWAS role-play task.

Results

Sample Characteristics

Group means and standard deviations on all measures[1] and CWAS scales at first testing are shown in Table 1.

Insert Table 1 about here

One way ANOVA's between groups yielded nonsignificant *F* values on all measures, indicating initial group comparability. Derived means resemble normative data reported elsewhere (e.g., Galassi *et al.*, 1974) and indicated that this sample was comparable to subject populations from other studies. Furthermore, the range of the sample's scores indicated that subjects varied from low to high assertive (cf., Galassi & Galassi, 1976; Schwartz & Gottmann, 1976).

Rater Reliabilities

CWAS coder agreement, based on independent ratings of 40% of the scored protocols, randomly selected for double coding, averaged .91 (range, .85 to .97 within scales). Observer reliabilities with the TBCL, based on checks throughout 20% of all testing sessions and calculated as the number of agreements divided by the number of

agreements plus disagreements, averaged .84 (range, .68 to .92 within categories).

Order Effects

The effect of test order was evaluated by examining differences between persons tested first with the CWAS versus persons tested first with the self-report measures. T-tests on all 12 variables for the two groups' initial session scores yielded no significant differences.

Indices of Temporal Stability

Pearson product-moment correlation coefficients were calculated between test and retest scores for all groups. The results, presented in Table 2, indicated highly significant

---

Insert Table 2 about here.

---

correlations ($p < .001$) for all of the variables over all retest intervals, with the following exceptions: coefficients over four weeks for both the global self-rating of assertion and latency were significant at $p < .01$ only, and coefficients over ten weeks for the global self-ratings of assertion and anxiety were both not significant. Coefficients of stability and significance levels for scales within the CWAS are also reported in Table 2.

Thirty of the 36 test-retest correlations (for the 12 variables over the 3 intervals) were not significantly affected by order. With the 6 that were, 5 (Latency, SRI,

FNE, and global self-rating of assertion over 1 week; Duration over 4 weeks) were significantly ($p < .05$) higher when the role-play measure preceded self-report testing, and one (SRIA over 4 weeks) was significantly ($p < .05$) higher when self-reporting was first.

Although test-retest correlations index predictability using a linear rule, they do not address an issue critical in outcome research: group changes over time. To evaluate the degree and direction of these changes, t-tests for correlated means were calculated between test and retest scores for each group on all variables. Results are shown in Table 3; of the 12 measures, only the CWAS, CRI, and TBCL did not change

---

Insert Table 3 about here

---

significantly over time for at least one of the retest intervals.

Convergent and Discriminant Validities

Intercorrelations among the 12 variables were computed separately for each retest group at each testing session. Across all 6 matrices, values were consistently of the same magnitude and direction with the exception of both global self-ratings, which displayed marked variability. Consequently, Table 4 presents intercorrelations for combined groups at

the initial assessment, exclusive of the global self-ratings.

Insert Table 4 about here

Convergent validity coefficients, reflecting the degree of association between measures of the same characteristic using different methods, were significant ($P < .001$) for all of the assertion measures. Discriminant validity is based on three progressively more stringent criteria: (1) whether correlations between a measure and other measures of the same characteristic using different methods exceed correlations between a measure and other measures of different characteristics using different methods; (2) whether correlations between a measure and other measures of the same characteristic using different methods exceed correlations between a measure and other measures of different characteristic using the same method, and (3) whether the pattern of relationships between different characteristics is replicated across measurement methods (Campbell & Fiske, 1959). All of the assertion measures met the first criterion. Only the CWAS met the second criterion, due to the strong correlations between the self-report measures of assertion and the self-report measures of anxiety. It was therefore not possible for any of the assertion measures to meet the third, most stringent criterion (Humphreys, 1960).

Subsidiary Analyses

Sources of Observed Anxiety Level Variance. In light of the newness of standardized role-play measures, analyses of this measurement method's properties would seem a priority. In particular, identifying instrument sources controlling observed anxiety level variance, and potentially subject discomfort, is important. A t-test for dependent means indicated that subjects displayed significantly less anxiety during the CWAS instructional prelude than they did during the first eight CWAS item presentations at both test ($t=4.19$, $df=119$, $p < .001$) and retest ($t=2.02$, $df=119$, $p < .05$) sessions. Moreover, subjects displayed significantly less anxiety during item presentations than they did while responding to items at both test ($t=4.3$[illegible], $df=119$, $p < .0001$) and retest ($t=2.07$, $df=119$, $p < .05$) sessions, despite the fact that the observation interval during responding ($\bar{X} \approx 3.6$ seconds) was far shorter than the observation interval during item presentation ($\bar{X} \approx 18$ seconds). It would seem, then, that at least with this sample and this measure, observed anxiety is more a function of responding to items than it is of simple exposure to the items or to the testing situation.

Subject Typologies. Several persons have noted the importance of differentiating between assertion and aggression clinically (cf., Alberti, 1977; Hollandsworth, 1977; Lange & Jakubowski, 1976), and this would seem an important attribute

of assertion measures as well. To explore the utilities of the CWAS, CRI, and CSES in making these distinctions, the following procedure was followed. Since the CWAS scoring procedure provides for designations of individual responses as instances of subculturally-judged aggression, assertion, or submission, subject types were identified by intercorrelating the number of aggressive, assertive, and submissive responses appearing within protocols. The resulting coefficients (r between assertion and aggression, .02; r between submission and aggression, -.47; r between submission and assertion, -.86) indicated that submissive responses were significantly more likely to be associated with aggressive responses than assertive ones ($t=128$, $df=117$, $p < .001$) and suggested four subject typologies: aggressive, assertive, submissive, and aggressive-submissive. Frequency distributions of the number of aggressive, assertive, and submissive responses within protocols were examined, and cut-off scores were set at points in each distribution which marked sharp drops in each response type's frequency. Subjects (N=77) who represented "pure" cases of each of the typologies were identified on the basis of these cut-off scores: ≥ 5 aggressive and ≤ 17 submissive responses, aggressive; ≥ 33 assertive and ≤ 4 aggressive responses, assertive; ≥ 18 submissive and ≤ 4 aggressive responses, submissive, and; ≥ 5 aggressive and ≥ 18 submissive responses, aggressive-submissive.[2]

For heuristic purposes, one-way ANOVAs were calculated between the four groups on all measures, including the anxiety ones. Subsequent analyses of variables yielding significant F ratios were conducted using Tukey's Honestly Significant Difference test. Measures yielding significant F ratios and their respective group means are presented in Table 5. CWAS total scores differentiated between all four

---

Insert Table 5 about here

---

groups; TBCL and CSES scores both differentiated between assertive and submissive subjects, but neither distinguished between assertive and aggressive ones. Although the TBCL aggressive-assertive comparison was not significant, inspection of the means suggests that there may be a curvilinear relationship between assertion level and observed anxiety level, a relationship which might help resolve some inconsistent findings between studies that have applied linear models to explore the relationship (cf., Gambrill & Rickey, 1975; MacDonald, 1978; Schwartz & Gottmann, 1977).

Discussion

Test-retest correlations for the CWAS, CRI, and CSES were highly significant over all retest intervals, as were the correlations for the anxiety instruments included for validation purposes.[3] Over the one and four week intervals,

coefficients for the assertion measures, excluding the global self-rating, were essentially identical, ranging from .88 to .81. Over the ten week interval, the correlation remained at .81 for the CSES, while it fell to an equivalently lower level for both the CWAS (r=.57) and the CRI (r=.54). On the second index of stability, there were no significant differences between mean test and retest scores with either the CWAS or the CRI over any of the retest intervals; mean CSES scores, however, tended to increase over time and, over the one and four week intervals, to increase significantly.

Lower test-retest reliabilities are usually interpreted as indications of instrument error; particularly with long retest intervals and subjects tested during developmental stages, however, lower reliabilities can also be interpreted as indications of instrument sensitivity to veridical behavior change (Anastasi, 1976). Consideration of the information provided by both indices of stability taken together gives some direction about which interpretation is more plausible for the lowered, ten week correlations with the CWAS and the CRI. Test-retest difference scores detect changes in an individual's score relative to the same individual's score at a different point in time. Correlation coefficients, in contrast, detect changes over time in an individual's score's standing relative to the scores of all other retested individuals. Significant mean difference scores in the absence

of lowered test-retest correlations, then, as was the case with the CSES, indicate that there was change over time, but that the change was relatively constant and unidirectional throughout the sample. Lower test-retest correlations in the absence of significant mean difference scores, however, as was the case with both the CWAS and the CRI, indicate that change was shown by a portion of the sample, rather than the sample as a whole, and that the change was neither constant nor unidirectional within the changing portion. The latter circumstances seem more plausible with this sample of individuals tested during their first semester of college and retested after a ten-week interval. Comparing information yielded by both stability indices for the CWAS and CRI, whose data converged, against information yielded by both stability indices for the CSES, then, suggests that in this instance, the lower CWAS and CRI correlations over the ten week interval would seem more plausibly interpreted as arising from instrument sensitivity to veridical change rather than instrument error.

Test-retest reliabilities for both global self-ratings were comparatively low from the outset and declined to nonsignificant levels by the longest retest interval. With one exception, their test-retest mean difference scores were all significant. Moreover, their relationships to each other and to all other variables showed marked fluctuations across

both times and groups. These data indicate, then, that global self-ratings of assertion (and anxiety) are neither valid nor reliable and thus are not tenable as either dependent variables in outcome research or criterion measures in assessment research.

Intercorrelations among the remaining measures included in the multitrait-multimethod matrix supported the convergent validities of the CWAS, CRI, and CSES. The CWAS received support for its discriminant validity, although the support might well have been less pronounced had there been standardized behavior sample measures for independent social skills available for inclusion. The data did not lend strong support for the discriminant validities of either self-report measure of assertion: both the CRI and the CSES were more strongly associated with at least one of the self-report measures of anxiety than they were with the behavior sample measure of assertion. Failures to demonstrate discriminant validity may arise from either overlap in measured constructs or high method variance relative to content saturation in multitrait-monomethod instruments. Which or what combination of these two produced the present result is indeterminant from these data, although the relationships observed between the behavior sample measures would favor the latter interpretation of the data from the self-reports.

While the three assertion instruments included in this study demonstrated convergent validity, there was clearly not complete redundancy among them. In the absence of continuing validation research including the proper external validity studies, which of the measures is or are more valid for given purposes will remain an unsettled issue. Therefore, it remains necessary to include multiple dependent variables in studies involving assertion and to conduct data analyses which take partial instrument redundancy into account (cf., Kaplan & Litrownik, 1977).

Reference Notes

1. MacDonald, M. L. The College Women's Assertion Sample. Copyrighted manual, 1978. (Available from Marian L. MacDonald, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003).
2. MacDonald, M. L. A Double Cross-Validation of the College Women's Assertion Sample. Manuscript submitted for publication, 1978.
3. Kern, J. M. & MacDonald, M. L. Context-Induced Differences in "Standardized" Measures' Scores: A Rose Is a Rose Is a Dandelion. Manuscript submitted for publication, 1978.

References

Alberti, R. E. Comments on "Differentiating assertion and aggression: Some behavioral guidelines." Behavior Therapy, 1977, 8, 353-354.

Anastasi, A. Psychological Testing (4th ed.). New York: Macmillan, 1976.

Compbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Eisler, R. M., Hersen, M., & Miller, P. M. Effects of modeling on components of assertive behavior. Journal of Behavior Therapy and Experimental Psychiatry, 1973, 4, 1-6.

Eisler, R. M., Miller, P. M., & Hersen, M. Components of assertive behavior. Journal of Clinical Psychology, 1973, 29, 295-299.

Endler, N. S., Hunt, J. McV., & Rosenstein, A. J. An S-R Inventory of Anxiousness. Psychological Monographs, 1962, 76, No. 536.

Friedman, P H. The effects of modeling and role-playing on assertive behavior. In R. D. Rubin, H. Fensterheim, A. A. Lazarus, & C. M. Franks (Eds.), Advances in Behavior Therapy. New York: Academic Press, 1971.

Galassi, J. P., Delo, J. S. Galassi, M. D., & Bastien, S. The College Self-Expression Scale: A measure of assertiveness. Behavior Therapy, 1974, 5, 165-171.

Galassi, M. D. & Galassi, J. P. The effects of role-playing variations on the assessment of assertive behavior. Behavior Therapy, 1976, 3, 343-347.

Galassi, J. P., Galassi, M. D., & Litz, M. C. Assertive training in groups using video feedback. Journal of Counseling Psychology, 1974, 21, 390-394.

Galassi, J. P., Hollandsworth, J. G., Jr., Radecki, J. C., Gay, M. L., Howe, M. R., & Evans, C. L. Behavioral performance in the validation of an assertiveness scale. Behavior Therapy, 1976, 7, 447-452.

Gambrill, E. D. & Rickey, C. A. An assertion inventory for use in assessment and research. Behavior Therapy, 1975, 6, 550-561.

Hersen, M. & Bellack, A. S. Assessment of social skills. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), Handbook of Behavioral Assessment. New York: Wiley, 1977.

Hersen, M., Eisler, R. M., & Miller, P. M. Development of assertive responses: Clinical, measurement, and research considerations. Behaviour Research and Therapy, 1973, 11, 505-521.

Hersen, M., Eisler, R. M., Miller, P. M., Johnson, M. B., & Pinkston, S. G. Effects of practice, instructions, and modeling on components of assertive behavior. Behaviour Research and Therapy, 1973, 11, 443-451.

Hollandsworth, J. G. Differentiating assertion and aggression: some behavioral guidelines. Behavior Therapy, 1977, 8, 347-352.

Humphreys, L. G. Note on the multitrait-multimethod matrix. Psychological Bulletin, 1960, 57, 86-88.

Kaplan, R. M. & Litrownik, A. J. Some statistical methods for the assessment of multiple outcome criteria in behavioral research. Behavior Therapy, 1977, 8, 383-392.

Kazdin, A. E. Covert modeling, imagery assessment, and assertive behavior. Journal of Consulting and Clinical Psychology, 1975, 43, 716-724.

Kern, J. M., & MacDonald, M. L. Assessing assertion: An investigation of construct validity and reliability. Journal of Consulting and Clinical Psychology, 1980, in press.

Lange, A. J. & Jakubowski, P. Responsible Assertive Behavior: Cognitive/Behavioral Procedures for Trainers. Champaign, Illinois; Research Press, 1976.

MacDonald, M. L. A behavioral assessment methodology applied to the measurement of assertion (Doctoral dissertation, University of Illinois at Champaign-Urbana, 1974). Dissertation Abstracts International, 1975, 35, 6101B. (University Microfilms No. 75-11, 819)

MacDonald, M. L. Measuring assertion: A model and method. Behavior Therapy, 1978, 8, in press.

MacDonald, M. L. & Tobias, L. L. Withdrawal causes relapse? Our response. Psychological Bulletin, 1976, 83, 448-451.

Mandler, G., Mandler, J. M., & Uviller, E. T. Autonomic feedback: The perception of autonomic activity. Journal of Abnormal and Social Psychology, 1958, 56, 367-373.

McFall, R. M. & Lillesand, D. B. Behavior rehearsal with modeling and coaching in assertion training. Journal of Abnormal Psychology, 1971, 77, 313-323.

McFall, R. M. & Marston, A. R. An experimental investigation of behavior rehearsal in assertion training. Journal of Abnormal Psychology, 1970, 76, 295-303.

McFall, R. M. & Twentyman, C. T. Four experiments on the relative contribution of rehearsal, modeling, and coaching to assertion training. Journal of Abnormal Psychology, 1973, 81, 199-218.

Nay, W. R. Analogue measures. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), Handbook of Behavioral Assessment, New York: Wiley, 1977.

Nietzel, M. T. & Bernstein, D. A. Effects of instructionally-mediated demand on the behavioral assessment of assertiveness. Journal of Consulting and Clinical Psychology, 1976, 44, 500.

Paul, G. L. Insight versus Desensitization in Psychotherapy. Stanford: Stanford University Press, 1966.

Richardson, F. C. & Tasto, D. L. Development and factor analysis of a social anxiety inventory. Behavior Therapy, 1976, 7, 453-462.

Schwartz, R. M. & Gottmann, J. M. Towards a task analysis of assertive behavior. Journal of Consulting and Clinical Psychology, 1976, 44, 910-920.

Taplin, P. S. & Reid, J. B. Effects of instructional set and experimenter influence on observer reliability. Child Development, 1973, 44, 547-554.

Watson, D. & Friend, R. Measurement of social-evaluative anxiety. Journal of Consulting and Clinical Psychology, 1972, 39, 246-252.

Footnotes

Requests for reprints may be sent to either author and should be addressed to the Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11794

[1]Since the three scores derivable from the CRI (assertion, nonassertion, and difference scores) were correlated consistently at .92 in this sample, results are reported throughout only for the assertion score.

[2]CSES and CRI means for the assertive and submissive groups identified through this procedure closely approximated group means in studies designed to explore differences between high and low assertive groups (cf., Galassi et al., 1976; Schwartz & Gottmann, 1976).

[3]It is important to note that reliabilities established in this work do not guarantee that scales will show similar reliabilities in other contexts: there is no such thing as the reliability of a scale (MacDonald & Tobias, 1976). Testing contexts are important considerations (cf., Kern & MacDonald, Note 3), and demand characteristics may alter indices of stability in ways not addressed by this study (cf., Nietzel & Bernstein, 1976).

Table 1

Group Means and Standard Deviations

for Measures and CWAS Scales at Initial Testing

	1 Week Retest Interval		4 Week Retest Interval		10 Week Retest Interval	
Variable	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
			Measures			
CWAS	5.52	.5	5.47	.5	5.48	.4
TBCL	55.28	17.8	54.20	17.7	53.20	25.2
Latency	2.65	2.7	2.26	1.8	2.47	1.2
Duration	3.18	1.9	3.59	2.0	3.14	1.9
CSES	120.45	18.2	125.95	17.5	125.83	18.6
CRI	13.80	5.6	15.72	5.7	15.45	5.0
Global Self-Rating: Assertion	4.28	1.3	4.77	1.2	4.25	1.2
FNE	17.58	4.1	15.93	8.2	15.55	6.9
SRI	27.38	5.9	25.28	5.9	26.13	5.6
SRIA	350.45	54.8	346.45	33.2	358.40	64.8
APQ	47.33	30.4	45.28	22.6	53.38	32.7
Global Self-Rating: Anxiety	5.05	1.1	4.75	1.5	4.83	1.0
			CWAS Scales			
I	5.67	.4	5.50	.5	5.46	.4
II	5.68	.8	5.46	.8	5.38	.7
III	5.02	1.2	4.93	1.5	5.01	1.1
IV	5.49	.8	5.14	.8	5.34	.8
V	6.30	1.1	5.78	.8	5.98	1.0
VI	6.52	.8	6.28	.9	6.07	1.0
IX	6.70	.8	6.36	1.0	6.27	.9

Note: Scores on Scale IX, intended as a scale to assess skill in detecting situations for which assertion is not appropriate, were significantly ($t=11.78$, $df=119$, $p < .001$; replicated within all groups separately at $p < .01$) higher and thus more submissive than total CWAS scores, a result providing some support for the scale's sensitivity (see MacDonald, 1974).

Table 2

Test-Retest Reliabilities for Measures and CWAS Scales Over 1 Week, 4 Week, and 10 Week Intervals

	1 Week Interval	4 Week Interval	10 Week Interval
Measures			
CWAS	.85	.84	.57
TBCL	.70	.76	.55
Latency	.94	.41**	.81
Duration	.80	.87	.86
CSES	.88	.88	.81
CRI	.81	.85	.54
Global Self-Rating: Assertion	.61	.41**	.24[a]
FNE	.84	.92	.79
SRI	.61	.84	.81
SRIA	.70	.80	.87
APQ	.83	.65	.65
Global Self-Rating: Anxiety	.65	.62	.13[a]
CWAS Scales			
I	.65	.75	.10[a]
II	.71	.83	.56
III	.69	.77	.57
IV	.64	.34*	.34*
V	.77	.45**	.44**
VI	.43**	.48	.31*
IX	.63	.69	.60

Note: All coefficients were significant at $p < .001$ except those so indicated.

[a]These correlations were not significant.

*$p < .05$

**$p < .01$

Table 3

Mean Test to Retest Difference Scores for Measures and CWAS Scales Over the Three Retest Intervals

	2 Week Interval	4 Week Interval	10 Week Interval
	Measures		
CWAS	.06	-.06	.04
TBCL	1.57	-.03	4.60
Latency	-.99**	-.84**	-.47
Duration	-.96**	-.64**	-.42
CSES	2.95*	3.33*	1.97
CRI	.03	.40	1.08
Global Self-Rating Assertion	.66**	-.20	.43*
FNE	-.63	-.95*	-2.90**
SRI	-1.35	-1.60*	-1.72**
SRIA	-11.65	-19.10**	-12.30*
APQ	-13.05**	-18.83**	-17.78**
Global Self-Rating: Anxiety	-.45**	-.30**	-.35**
	CWAS Scales		
I	.07	-.00	.12
II	.00	-.02	.05
III	.33	.06	-.04
IV	.04	-.23	.18
V	.01	-.13	-.08
VI	.06	-.08	-.40
IX	.18	.02	.03

Note: Positive and negative scores indicate increased and decreased scores, respectively, on the measured characteristic from test to retest.

*$p < .05$

**$p < .01$

Table 4

Intercorrelations of All Variables for the Combined Groups at the Initial Assessment

Instruments Identified by "Trait"	Method I (Behavior Sample) Instruments				Method II (Self-Report) Instruments					
	CWAS	TBCL	Latency	Duration	CSES	CRI	FNE	SRI	SRIA	APQ
CWAS: Assertion	1.00									
TBCL: Anxiety	-.19*	1.00								
Latency: Anxiety	-.21*	.00	1.00							
Duration: Anxiety	.16*	.20*	.09	1.00						
CSES: Assertion	.49***	.03	-.07	.02	1.00					
CRI: Assertion	.28***	.03	-.02	.13	.51***	1.00				
FNE: Anxiety	-.18*	.07	-.11	.10	-.62***	-.37***	1.00			
SRI: Anxiety	-.05	-.03	-.11	.05	-.57***	-.27***	.59***	1.00		
SRIA: Anxiety	.04	-.09	-.14	.04	-.31***	-.27***	.40***	.43***	1.00	
APQ: Anxiety	.20*	-.02	-.08	.20*	-.23**	-.16*	.39***	.33***	.63***	1.00

*$p < .05$

**$p < .01$

***$p < .001$

Table 5

Means, F Ratios, and Group Distinctions for Measures Successfully Differentiating Between Subject Types

	ANOVA Value			Subject Type			
Measure	F	df	p	Aggressive	Assertive	Submissive	Aggressive-Submissive
CWAS	82.73	(3,73)	>.00001	4.98a	5.26b	6.18c	5.78d
TBCL	3.06	(3,73)	>.033	53.22ab	43.96a	68.15b	59.17ab
CSES	9.04	(3,73)	>.0001	135.60a	127.96a	112.85b	119.33ab
APQ	5.20	(3,73)	>.003	44.78a	58.41a	37.34a	35.00a

Note: Means in the same row with different subscripts are significantly ($p < .05$) different by Tukey's Honestly Significant Difference test; higher CWAS scores indicate greater submission.